Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

COMPUTING WORST-CASE TAIL PROBABILITIES IN CREDIT RISK

Soumyadip Ghosh

IBM T. J. Watson Research Center
Yorktown Heights, NY 10598, U.S.A.

Sandeep Juneja

Tata Institute of Fundamental Research
Homi Bhabha Road, Colaba
Mumbai, MH 400005, INDIA

ABSTRACT

Simulation is widely used to measure credit risk in portfolios of loans, bonds, and other instruments subject to possible default. This analysis requires performing the difficult modeling task of capturing the dependence between obligors adequately. Current methods assume a form for the joint distribution of the obligors and match its parameters to given dependence specifications, usually correlations. The *value-at-risk* risk measure (a function of its tail quantiles) is then evaluated. This procedure is naturally limited by the form assumed, and might not approximate well the "worst-case" possible over *all* joint distributions that match the given specification. We propose a procedure that approximates the joint distribution with *chessboard* distributions, and provides a sequence of improving estimates that asymptotically approach this "worst-case" value-at-risk. We use it to experimentally compare the quality of the estimates provided by the earlier procedures.

1 INTRODUCTION

A remarkable increase in interest has been observed in the past few years in methods that analyse credit related risk. An accurate knowledge of the risk involved in the credit portfolio of a financial institution helps it understand and choose an appropriate risk trading, hedging and transferring strategy. Developments in banking supervisions, for instance the Basel accords (Basel Committee 2002), now require better analysis and management of credit risk.

Monte Carlo simulation is amongst the most widely used risk assessment tools. The basic principle remains the same as in other applications: a number of possible scenarios of the values of the portfolio are sampled from an appropriately chosen model and the desired risk measure, most commonly the *value-at-risk*, is then calculated. The value-at-risk measure captures the likelihood of multiple defaults occurrences resulting in large losses (Section 1.1 provides a formal definition).

Portfolios consist typically of a large ($\sim 200 - 250$) collection of obligors who have been extended credit in some form or the other. One of the fundamental problems in this context is that of modeling the dependence within a portfolio adequately. A popular framework used to model the risk of a credit portfolio assumes that the portfolio value is driven by latent variables that represent individual obligor credit values. The dependence between the obligors can be captured in their joint distribution, for instance by representing the latent variables in terms of common factors like various macro-economic or industry-specific effects. The *normal copula model* which assumes that the obligor credit values follow a multivariate normal distribution is one of the most widely used models in practice. It is the basis of the Importance-Sampling based risk assessment methods constructed in Glasserman and Li (2005), and finds application in many popular risk management tools such as J.P. Morgan's CreditMetrics (Gupta et al. 1997) and Moody's KMV system (Kealhofer and Bohn 2001).

Recent empirical work has shown that in certain instances a significant co-movement of obligor values can be observed (Mashal and Zeevi 2003). This leads Glasserman et al. (2002) and Bassamboo et al. (2007) to consider joint distributions with fatter tails that allow *extremal dependence* (intuitively, this means that obligors can take on large losses with non-negligible probability). These include the multivariate t-distribution.

The distributional forms assumed above are customized for a specific instance of a portfolio by finding appropriate values for its parameters that match certain dependence measures. The dependency is often assumed given as a set of pairwise correlation values. These methods can thus be considered within a common framework where a risk measure (which is a function of the joint distribution of the obligors) is evaluated or estimated for different joint distributions that match the same set of specifications, namely *marginal* distributions of each obligors and a matrix of pairwise correlations.

1-4244-0501-7/06/$20.00 ©2006 IEEE

The normal- and t-copula based methods can thus produce risk estimates that differ for the same base marginals and correlations. Choosing between these estimates can be tricky since the nature of the difference will depend on the risk measure and correlation values. Our contention here is that in a risk analysis setting knowledge of the "worst-case" is more important, and these estimates may not be close to the worst-case value of the risk measure. This is because the strong structural assumptions made by these methods prevent them from fully "maximizing" the impact of the risk.

We present an alternate approach that drops all structural assumptions on the joint distribution beyond the marginals and second moments. Section 1.1 gives the formal statement of the *value-at-risk* problem we study, and Section 1.2 discusses our contributions.

1.1 Formal Definition

A financial institution maintains a portfolio of credits it has extended to d obligors through various financial instruments that include loans, corporate bonds, etc. We are interested in evaluating the chance that the loan portfolio defaults after a preset time period, that is, its value falls below a threshold. Let X_i represent the value of the i^{th} obligor at maturity, and F_i its distribution, henceforth termed its *marginal* distribution. We model the joint behaviour of the obligors by matching the *joint distribution* of the random vector $X = (X_1, \ldots, X_d)$ with the given set of marginals and a *covariance matrix* Σ. The (i, j)th element of this matrix represents the Pearson product-moment covariance between the two components X_i and X_j, defined when $E(X_i^2 + X_j^2) < \infty$, and is given by

$$\Sigma(i, j) = \mathrm{Cov}(X_i, X_j) = EX_iX_j - EX_iEX_j.$$

The i^{th} obligor individually defaults if the value of her instrument is below the threshold v_i, where v_i is the p_i^{th} quantile of F_i, i.e., $F_i(v_i) = p_i$, $\forall i = 1, \ldots, d$. The values p_i are pre-determined or fixed. Let e_i represent the loss incurred by the portfolio manager if obligor i defaults. This may be a discounted value of the loan provided her. We are interested in obtaining an accurate estimate of the value-at-risk quantity

$$R(u) \quad \triangleq \quad \max_{\mu \in \mathcal{F}} \quad \mu\,(L > u)\,, \tag{1}$$

where

$$L \quad \triangleq \quad \sum_{1}^{d} e_i I(X_i \leq v_i), \text{ the } \textit{loss} \text{ function} \tag{2}$$

$$\begin{aligned} \mathcal{F} \quad &\triangleq \quad \{ \text{ all probability measures } \mu : \\ &\qquad \mu_{(i)} = F_i, \forall i = 1, \ldots, d, \\ &\qquad \text{and } \mathrm{Cov}(\mu) = \Sigma \quad \}, \quad \text{and} \\ I(A) \quad &= \quad 1 \text{ if event A is true, and } 0 \text{ otherwise.} \end{aligned} \tag{3}$$

The set $\mathcal{F}$ is closed; the argument in the proof of Theorem 3.1 can be modified to provide this result. Hence the maximum in (1) makes sense.

1.2 Contributions

Our approach models the joint distribution of the obligor-values X_i as a *chessboard distribution*. Section 2 gives a brief primer on this class of distributions. A sequence of specially formulated linear programs are solved to obtain estimates for the $R(u)$. The estimation procedure is described in Section 3. The primary result of interest (Theorem 3.1) shows that as the size of the linear program gets large, the estimate values asymptotically approaches R. Thus, this procedure provides estimates that approach the true maximal value of $R(u)$ as defined in (1).

Our primary aim in this study is to show that structural assumptions on the joint distributions can have a pronounced effect on the value-at-risk approximations. Towards this end, we perform the numerical experiments described in Section 6. These compare the estimates from the earlier methods against those obtained from our procedure to see how accurate the methods are in providing estimates of the quantity R. The results of these tests will be presented at the conference.

This procedure can conceivably be used as an alternate method to estimate the value-at-risk $R(u)$. The approximating linear programs (Section 3) however get harder to solve with the parameter d since the sizes of these programs grow very quickly with d. So, while credit-portfolios consist typically of hundreds of obligors, solving them exactly for values higher than say $d \geq 20$ may not be advisable. However, the chessboard LPs have a very specific structure which can be exploited to obtain approximations of high quality. Specifically they are also known as *axially symmetric multi-index transportation problems (A-MITPs)* with additional constraints (Queyranne and Spieksma 2001). Efficient approximation algorithms are available for A-MITPs (Queyranne and Spieksma 1997), and extending these to handle additional constraints is an avenue being actively pursued by the authors.

Section 4 discusses how another aspect of risk analysis of large-portfolios can provide some respite when applying our proposed procedure. Essentially, large portfolios are assumed to consist of groups of homogeneous obligors (defined in Section 4). The structure of chessboard distributions then allows us a simplification that can roughly be thought of as replacing an entire group of homogeneous obligors

with a single representative, thus reducing the overall dimension of the chessboard LP. This simplification is tested in the results in Section 6.

2 CHESSBOARD DISTRIBUTIONS

Chessboard distributions were introduced in Ghosh and Henderson (2002) where they were used to determine whether a specified product-moment covariance matrix is feasible for a given set of marginals. We shall illustrate the construction of chessboard distributions for a $3-$dimensional random vector $X = (X_1, X_2, X_3)$ with marginal distributions that have densities. These requirements are only for notational convenience; the results can be extended to include all marginal distributions with finite second moments, and extend transparently to higher dimensions.

Suppose we wish to construct a chessboard distribution for X such that each of the marginals has a density f_i and finite variance. Furthermore, let domain $\text{dom}(F_i)$ of the marginal F_i be the entire real line (the method is easily tailored for semi-infinite or finite $\text{dom}(F_i)$).

Let

$$\{y^n_{i,j_i} : i = 1,2,3,\ j_i = 1,\ldots,m(n)\}$$

be a set of points (not including $\pm\infty$) that divide the $\text{dom}(F_i)$ into $m(n)$ sub-intervals. We shall occasionally drop the super-script n when the contexts allows us the freedom. Let M_i^- and M_i^+ represent the leftmost and rightmost finite points respectively. Thus, if X_i were exponentially distributed, $M_i^- = y_{i,1} = 0$, and $M_i^+ = y_{i,m(n)}$.

The range can be divided in any manner, as long as the points satisfy two conditions. First, the internal mesh becomes dense, i.e.,

$$\lim_{n\to\infty} \sup_{i,j_i} |y^n_{i,j_i} - y^n_{i,j_i-1}| = 0. \tag{4}$$

Second, the end-points should satisfy

$$\min_i |M_i^{\pm}| \to \infty \text{ as } n \to \infty. \tag{5}$$

These conditions are satisfied for the general case of marginals with $\text{dom}(F_i) = (\infty, +\infty)$ by a choice of grid-points

$$\begin{aligned} \mathcal{G}(n) \quad = \quad & \{y^n_{i,j_i} = -\frac{1}{2}\sqrt{n} + \frac{j_i - 1}{\sqrt{n}} : \\ & i = 1,2,3,\ j_i = 1,\ldots,n+1\}. \end{aligned} \tag{6}$$

Here, $|M_i^+| = |M_i^-| = \frac{1}{2}\sqrt{n}$. In the sequel, we will work on the grid is $\mathcal{G}(n)$. Let $y(v)$ be the grid-point nearest to the point $v = (v_1, \ldots, v_d)$ of individual default thresholds. We shall perturb the grid by $(v - y(v))$ so that the grid contains the point v. In what follows, we shall then assume that $\mathcal{G}(n)$ represents the grid defined in (6) perturbed by $(v - y(v))$.

For $1 \le j_1, j_2, j_3 \le n+1$ define *cell* $C(j_1, j_2, j_3)$ to be the (j_1, j_2, j_3)th rectangular region

$$\{x = (x_1, x_2, x_3) : y_{i,j_i-1} < x_i \le y_{i,j_i},\ i = 1,2,3\} \cap \Re^3.$$

Let $q(j_1, j_2, j_3) = P(X \in C(j_1, j_2, j_3))$ to be the probability that the constructed random vector appears in the (j_1, j_2, j_3)th cell. The chessboard distribution is defined so that within each cell the components of X are independent and distributed according to the desired marginals f_1, f_2, f_3 restricted to the cell $C(j_1, j_2, j_3)$. Let $p_{i,k} = P(X_i \in (y_{i,k-1}, y_{i,k}])$ be the probability that the ith marginal random variable lies in the kth sub-interval. The density $f(x)$ of X evaluated at $x \in C(j_1, j_2, j_3)$ is then given by

$$q(j_1, j_2, j_3) \frac{f_1(x_1)}{p_{1,j_1}} \frac{f_2(x_2)}{p_{2,j_2}} \frac{f_3(x_3)}{p_{3,j_3}}. \tag{7}$$

To be consistent with the given marginals, the $q(j_1, j_2, j_3)$ values must satisfy the constraints

$$\begin{aligned} \sum_{j_2,j_3=1}^{n} q(j_1, j_2, j_3) &= p_{1,j_1}, \quad j_1 = 1,\ldots,n+1 \\ \sum_{j_1,j_3=1}^{n} q(j_1, j_2, j_3) &= p_{2,j_2}, \quad j_2 = 1,\ldots,n+1 \\ \sum_{j_1,j_2=1}^{n} q(j_1, j_2, j_3) &= p_{3,j_3}, \quad j_3 = 1,\ldots,n+1 \\ q(j_1, j_2, j_3) &\ge 0 \qquad 1 \le j_1, j_2, j_3 \le n+1. \end{aligned} \tag{8}$$

Theorem 2.3.1 in Ghosh (2004) asserts the following.

Theorem 2.1 *If q satisfies the constraints (8), and X is constructed with the chessboard density f defined in (7), then X has the desired marginals F_i.*

Ghosh and Henderson (2002) give a procedure to find a chessboard distribution that matches desired values Σ of the product-moment covariance matrix Σ^{ch} of the chessboard density (7). Chessboards, by construction, have the right marginals, and the diagonal elements of the covariance matrices Σ and Σ^{ch} are determined by the marginal distributions. So only the upper diagonal elements of the (symmetric) correlation matrix have to be matched. Matching correlations using chessboards is advantageous because the correlations $\Sigma^{ch}(i,j)$ induced by the chessboard density can be written as a linear function of the $q(\cdot)$s. Specifically,

$$\Sigma^{ch}(1,2) = \sum_{j_1,j_2,j_3} \gamma_{1,j_1} \gamma_{2,j_2} q(j_1, j_2, j_3) - EX_1 EX_2, \tag{9}$$

where, for $1 \le i \le 3$ and $1 \le m \le n$,

$$\gamma_{i,m} = E[X_i | X_i \in (y_{i,m-1}, y_{im}]] \tag{10}$$

is the conditional mean of X_i given that it lies in the mth sub-interval (which is determined by its marginal density f_i). Thus the differences $|\Sigma^{ch}(i,j) - \Sigma(i,j)|$ too are linear functionals of the $q(\cdot)$s. The correlations can be matched by minimizing the difference $r(\Sigma^{ch}, \Sigma)$, subject to the constraints (7), where

$$r(\Sigma^{ch}, \Sigma) = \sum_{1 \le i < j \le 3} \left| \Sigma^{ch}(i,j) - \Sigma(i,j) \right|. \tag{11}$$

These linear programs can be augmented with upper bounds on the value of the objective function via bounds of the form

$$\left| \Sigma^{ch}(i,j) - \Sigma(i,j) \right| \le B^n(i,j). \tag{12}$$

Ghosh (2004) show that bounds $B(i,j)$ can be constructed such that with great generality they obey $B^n(i,j) \to 0$ as $n \to \infty$.

Ghosh and Henderson (2002) use chessboard LPs augmented with the (12) bounds to investigate the feasibility of correlation matrices. A correlation matrix is *feasible* for a given set of marginals if a joint distribution function exists with these marginals and correlation values. The augmented LPs help Ghosh (2004) show in Theorem 2.3.5 that for almost any (in a precise probabilistic sense) feasible covariance matrix Σ there exists a finite n such that the corresponding linear program (11) augmented with the bounds (12) has an optimal value of 0. Additionally, Theorem 2.3.2 in Ghosh (2004) gives that a covariance matrix is infeasible for the given marginals if, and only if, the augmented chessboard LP is infeasible for some $n \ge 1$.

Our objective here is to maximize the value-at-risk probability in (1). We shall provide in Section 3 a linear program formulation to estimate $R(u)$ that shall use the density function equations (8) and the correlation bounds (12). But first we provide some additional information on these bounds and a proposition we shall require for the proof of the asymptotic consistency in Theorem 3.1.

2.1 Bounds on the Covariances

Bounds $B^n(i,j)$ are derived by first supposing that there exists a random vector $\tilde{X}$ with the prescribed covariance matrix Σ. We can then redistribute the probability mass of its distribution within cells (thus keeping the cell probability masses constant) so that the conditional density given a cell is one of independent random variables with the desired marginals, that is, it follows the chessboard pattern. Let X denote a random vector with the redistributed probability mass. We provide a bound on the change in covariance due to this redistribution. This gives the bound (12).

We have for $\mathcal{G}(n)$ that $|M_i^-| = |M_i^+| \stackrel{\triangle}{=} M_i = \frac{1}{2}\sqrt{n}$. Let $\mathcal{C}(n)$ represent the the part of the support of X (and $\tilde{X}$) bounded by the rectangle $[-M_1, M_1] \times [-M_2, M_2]$. As per our notation, $\mathcal{C} = \mathcal{C}(n)$ is given by the collection of cells $C(j_1, j_2, j_3)$ with indices j_1 and j_2 ranging over $2, \ldots, n-1$. The absolute change in covariance due to the redistribution operation, $|EX_1X_2 - E\tilde{X}_1\tilde{X}_2|$, is bounded above by the sum:

$$\le \left| E[X_1X_2 - \tilde{X}_1\tilde{X}_2 | \mathcal{C}\}] \right| \tag{13}$$

$$+ \quad \left| E[X_1X_2 - \tilde{X}_1\tilde{X}_2 | \mathcal{C}^c\}] \right|. \tag{14}$$

Ghosh (2004) show how the second term (14), which includes cells of infinite length, can be bounded by terms that approach 0 as $n \to \infty$. For the proof of the chief result Proposition 2.2 in this section, it will be sufficient to note that this bound has the form

$$\le \quad 2 \quad E[X_1^2 \, I\{|X_1| > \frac{1}{2}\sqrt{n}\}\,]^{1/2} E[X_2^2]^{1/2} \tag{15}$$

$$+ \quad 2 \quad E[X_2^2 \, I\{|X_2| > \frac{1}{2}\sqrt{n}\}\,]^{1/2} E[X_1^2]^{1/2}.$$

Since the variances of all components of X are finite, the bound approaches 0 as $n \to \infty$. The bound is independent of the $q(\cdot)$s, and the rate at which it vanishes depends on the tail behaviour of the marginal distributions.

The first term (13) represents the change due to the redistribution operation in a compact part $\mathcal{C}$ of the support. This can be strongly bounded in a fashion similar to that used in Ghosh and Henderson (2002). For the compact region, observe that

$$E[X_1X_2 - E\tilde{X}_1\tilde{X}_2 | \mathcal{C}] = \tag{16}$$

$$\sum_{j_1,j_2,j_3=1}^{n-1} \left\{ (\gamma_{1,j_1}\gamma_{2,j_2} - E[\tilde{X}_1\tilde{X}_2 | \tilde{X} \in C(j_1,j_2,j_3)] \right\} \times \quad q(j_1,j_2,j_3),$$

where $\gamma_{i,m}$ is defined as in (10). But

$$y_{1,j_1-1}\, y_{2,j_2-1} \le E[\tilde{X}_1\tilde{X}_2 | \tilde{X} \in C(j_1,j_2,j_3)] \le y_{1,j_1}\, y_{2,j_2}. \tag{17}$$

Combining (16) with (17) we get the bounds

$$E[X_1X_2 - E\tilde{X}_1\tilde{X}_2 | \mathcal{C}] \tag{18}$$

$$\le \quad \sum_{j_1,j_2,j_3=1}^{n} \tilde{q}(j_1,j_2,j_3)(\gamma^n_{1,j_1}\,\gamma^n_{2,j_2} - y^n_{1,j_1-1}\, y^n_{2,j_2-1})$$

$$\ge \quad \sum_{j_1,j_2,j_3=1}^{n} \tilde{q}(j_1,j_2,j_3)(\gamma^n_{1,j_1}\,\gamma^n_{2,j_2} - y^n_{1,j_1}\, y^n_{2,j_2}).$$

The super-scripts n have been included to emphasize the dependence on the discretization parameter n. Now, for the grid $\mathcal{G}(n)$, $\gamma^n_{i,j_i} - y^n_{i,j_i-1} \leq \frac{1}{\sqrt{n}}$. An easy calculation yields that the upper bound (18) is of $O(\frac{1}{\sqrt{n}})$, and a similar result holds for the lower bound. In general, the terms on the right hand of the bounds in (18) are $O(|y^n_{i,j_i} - y^n_{i,j_i-1}|)$. Assumption (4) then ensures that these approach 0 as $n \to \infty$.

Putting together the bound (18) with (13), and the bound in (15) with (14) then gives us a bound $B^n(i,j)$ on the difference $|\Sigma^{ch}(i,j) - \Sigma(i,j)|$. By the properties observed above, the rate at which $B^n(i,j)$ approaches 0 depends on the rates of the convergences in (4) and (15), and is equal to the slower of the two. For instance, if the marginal distributions have an exponential tail, then the terms from (15) also decay exponentially, and the correlations that satisfy (12) differ from Σ by no more than $O(\frac{1}{\sqrt{n}})$. On the other hand, if the marginal distributions have a fat tail, for instance they decay at a rate $\sim x^{-(2+\varepsilon)}$, then the terms in (15) dominate, with a decay rate of $n^{-\varepsilon/2}$.

The preliminary result described next will be needed in the main result Theorem 3.1. For each $n \geq 1$, let $\mathcal{M}(n,\Sigma)$ be the collection of all chessboard density functions constructed on grid $\mathcal{G}(n)$ that match the given marginals and satisfy the bounds (12) for a target correlation matrix Σ.

Proposition 2.2 *Suppose the mass of a chessboard density $\mu^{4n} \in \mathcal{M}(4n,\Sigma)$ (defined over $\mathcal{G}(4n)$) is re-distributed as described earlier in this section over the grid $\mathcal{G}(n)$ to form a chessboard μ^n. Then, $\mu^n \in \mathcal{M}(n,\Sigma)$.*

Thus, a chessboard on a finer grid $\mathcal{G}(4n)$ that satisfies the tighter constraint on its correlation (but does not exactly match Σ) can be re-distributed over a coarser grid $\mathcal{G}(n)$ such that the new distribution satisfies the corresponding relaxed bound.

Proof: Let q^n and Σ^n represent the chessboard parameters and covariance matrix of μ^n, and Σ^{4n} the covariance matrix corresponding to μ^{4n}. We need to show that the covariance differences $\Sigma^n(i,j) - \Sigma(i,j)$ are bounded above by the values in (12). We will show this is true for the $(1,2)-$element; the proof for the other elements is identical. Condition, as before, on parts $\mathcal{C}(n)$ and $\mathcal{C}^c(n)$ of the grid $\mathcal{G}(n)$ to get

$$\begin{aligned} |\Sigma^n(1,2) - \Sigma(1,2)| &\leq \{|\Sigma^n(1,2) - \Sigma(1,2)|\,|\mathcal{C}(n)\} \\ &+ \{|\Sigma^n(1,2) - \Sigma(1,2)|\,|\mathcal{C}^c(n)\}. \end{aligned}$$

The part conditioned on $\mathcal{C}^c(n)$ is bounded above by the terms in (15). These depend only on the marginal distributions of μ^n, and hence by construction the distribution μ^n matches the $\mathcal{C}^c(n)$ part of $B^n(i,j)$ in (12). Split the difference in the $\mathcal{C}(n)$ term into

$$\begin{aligned} &\{\Sigma^n(1,2) - \Sigma(1,2)|\mathcal{C}(n)\} \\ &\leq \left\{\Sigma^n(1,2) - \Sigma^{4n}(1,2)|\mathcal{C}((n)\right\} \\ &+ \left\{\Sigma^{4n}(1,2) - \Sigma(1,2)|\mathcal{C}(n)\right\}. \end{aligned}$$

The first term in this split evaluates to

$$\begin{aligned} &E[X^n_1 X^n_2 - EX^{4n}_1 X^{4n}_2 | \mathcal{C}(n)] = \\ &\sum_{j_1,j_2,j_3=1}^{n-1} \Big\{ (\gamma^n_{1,j_1}\gamma^n_{2,j_2} - E[X^{4n}_1 X^{4n}_2 | X^{4n} \\ &\qquad \in C^n(j_1,j_2,j_3)] \Big\} \quad \times \quad q^n(j_1,j_2,j_3). \end{aligned}$$

The compact region $\mathcal{C}(n)$ is contained within the compact region $\mathcal{C}(4n)$ of grid $\mathcal{G}(4n)$, and the cells of $\mathcal{G}(4n)$ fit neatly within the cells of $\mathcal{G}(n)$ with sides exactly half in length. Moreover since μ^{4n} is a chessboard distribution on $\mathcal{G}(4n)$, the conditional moments of the distribution μ^{4n} on the cell $C^n(j_1,j_2,j_3)$ of grid $\mathcal{G}(n)$ satisfy

$$\begin{aligned} E[X^{4n}_1 X^{4n}_2 | X^{4n} \in C^n(j_1,j_2,j_3)] &\geq \gamma^{4n}_{1,2j_1-1}\,\gamma^{4n}_{2,2j_2-1} \\ &\leq \gamma^{4n}_{1,2j_1}\,\gamma^{4n}_{2,2j_2}. \quad (19) \end{aligned}$$

The distribution μ^{4n} also satisfies the $4n-$version of the upper bound (18) because of its membership in $\mathcal{M}(4n)$. Combining (19) with this $4n$-version of$B^{4n}(1,2)$ in (18) we see that the expression $\{\Sigma^n(1,2) - \Sigma(1,2)|\mathcal{C}(n)\}$ can indeed be bounded by the n-version $B^n(1,2)$. An analogous proof for the lower bound thus gives us the result. □

3 A CHESSBOARD FORMULATION

The theory developed for the chessboard distributions can be modified for our objective of estimating the maximal value-at-risk probability $R(u)$ in (1). Split the region $\mathbf{R}^d$ into 2^d rectilinear sets S_t aligned to the standard axes and centered at $v = (v_i,\ldots,v_d)$. (Recall that the threshold v_i is the value below which the obligor i defaults.) The parts S_t are indexed by

$$t = \{t_i = I(x_i \leq v_i), \forall i = 1,\ldots,d,\}. \quad (20)$$

Let $\mathcal{S}(u) = \{S_t : \sum_1^d e_i I(x_i \leq v_i) > u \quad \forall x \in S_t\}$. The union $\mathcal{S}(u)$ is the region of interest in evaluating the function L in (1). Observe that since v is a grid point in $\mathcal{G}(n)$, each rectilinear set S_t and the region $\mathcal{S}(u)$ consists of a collection of cells $C^n(j_1,j_2,j_3)$ each included in its entirety. Denote the collection of cells in S_t as $J^n_t(u) = \{(j_1,j_2,j_3) : C^n(j_1,j_2,j_3) \subseteq \mathcal{S}_t(u)\}$, and the collection $J^n(u) = \cup_{t:S_t \subseteq \mathcal{S}} J^n_t(u)$. We solve a sequence of linear

programs that find

$$R^n(u) \quad = \quad \max_q \sum_{J^n(u)} q^n(j_1, j_2, j_3) \qquad (21)$$

such that q^n's satisfy the constraints (8) and the $n-$version covariance bounds in (12). Stated in other terms, the objective function (21) is maximized over the feasible set $q^n \in \mathcal{M}(n, \Sigma)$. The chief advantage of this formulation is that the estimates produced approach the true maximum $R(u)$ as n grows. Specifically, we have the following.

Theorem 3.1 *Suppose Σ is a feasible covariance matrix. Then the sequence of optimal objective values $\{R^n(u), n \geq 1\}$ of the LPs (21) obey*

$$R^n(u) \to R(u) \text{ as } n \to 0.$$

The feasibility of the correlation matrix Σ can be determined by the chessboard procedure of Ghosh (2004) in Chapter 2.

Proof of Theorem 3.1: Since Σ is assumed feasible, there exists a chessboard density of finite n_0 that matches the matrix Σ exactly(Ghosh and Henderson 2002). For any large $N \geq n_0$, we shall show that the optimal objective value $R^n(u) \to R(u)$ along the sub-sequence $\{n_k = 4^k N, k = 0, 1, 2, \ldots\}$. Every $n \geq n_0$ belongs to such a sub-sequence (either as a 4^k-multiple or as a seed N for such a sequence), and thus any arbitrary sub-sequence of $\{R^n(u), n \geq n_0\}$ has $R(u)$ as its limit. This, along with the fact that the sequence is bounded (trivially by $\{0, 1\}$), leads us to conclude that $R^n(u) \to R(u)$ as $n \to \infty$.

Consider the fixed sub-sequence $\{n_k\}$ defined above. Let μ^{n_k} represent the chessboard density determined optimal for the LP (21) of size n_k. Each μ^{n_k} is the distribution of a random vector with marginals possessing finite second moments. Hence, the sequence $\{\mu^{n_k} : k \geq 1\}$ is tight, and by Theorem 29.3 on p. 392 of Billingsley (1986), it possesses a weakly convergent sub-sequence $\{\mu^{n_k(l)} : l \geq 1\}$, converging to μ say.

Now, μ has the right marginals. This follows from the Mapping theorem (Theorem 29.2, p. 391 of Billingsley 1986) since each $\mu^{n_k(l)}$ has the marginals we desire, $\mu^{n_k(l)} \Rightarrow \mu$ as $k \to \infty$, and the projection map $\pi_j : \mathbf{R}^{d(d-1)/2} \to \mathbf{R}$ that returns the jth coordinate of a vector in $\mathbf{R}^{d(d-1)/2}$ is continuous. The marginals are assumed to possess a density. This means that no point masses are allowed in the marginals and so no mass is placed on the boundary of the axis-aligned rectilinear region $\mathcal{S}(u)$. The set $\mathcal{S}(u)$ is then $\mu-$continuous, and

$$\lim_{l \to \infty} R^{n_k(l)}(u) \; = \; \lim_{l \to \infty} \quad \mu^{n_k(l)}(\mathcal{S}(u)) \; = \; \mu(\mathcal{S}(u)). \quad (22)$$

If C^{n_k} is the covariance matrix of the distribution μ^{n_k}, then

$$\sum_{i=1}^{d-1} \sum_{j=i+1}^{d} |C_{ij}^{n_k} - \Sigma_{ij}| \to 0$$

as $k \to \infty$. This follows from the bounds $B(i, j)$ described in the preceding section.

Finally, if $X^{n_k(l)}$ has distribution $\mu^{n_k(l)}$, then $(X_i^{n_k(l)} X_j^{n_k(l)} : l \geq 1)$ is uniformly integrable. To see this, note that (let $m = n_k(l)$)

$$\begin{aligned} & \sup_m E\left[|X_1^m X_2^m I\{|X_1^m X_2^m| > K\}|\right] \\ & \leq \quad \sup_m E\left[|X_1^m X_2^m| I\left\{|X_1^m| > \sqrt{K}\right\}\right. \\ & + \quad \left. |X_1^m X_2^m| I\left\{|X_2^m| > \sqrt{K}\right\}\right] \end{aligned}$$

This holds because, for any two positive numbers x and y, $\{xy > K\} \subseteq \{\max\{x, y\} > \sqrt{K}\} \subseteq \{x > \sqrt{K}\} \cup \{y > \sqrt{K}\}$. An argument along the lines of those given in Section 2.1 in the context of the bounding expression in (15) shows that the expression on the right side converges to 0 as $K \to \infty$. This establishes the uniform integrability result. It immediately follows (Theorem 25.12 in Billingsley 1995) that the covariance matrix Λ of μ is given by

$$\Lambda = \lim_{k \to \infty} C^{n_k} = \Sigma.$$

Thus, μ has the required marginals and covariance matrix. In other words, $\mu \in \mathcal{F}$, where $\mathcal{F}$ is defined as in (3).

For any distribution in $\mathcal{F}$ that achieves the maximum $R(u)$, observe that this distribution can be rearranged (in the fashion described in Section 2.1 to derive the bounds (12)) to obtain a feasible solution for each of the LPs (21) of size n. Thus, the optimal values $R^{n_k(l)}(u) \geq R(u)\ \forall l$, and this in conjunction with (22) and the fact that the limit $\mu \in \mathcal{F}$, gives us that $R^{n_k(l)}(u) \to R(u) = \mu(T(u))$.

Proposition 2.2 tells us that the optimal chessboard found over the grid $\mathcal{G}(4n)$ is a feasible solution to the LP (21) over the grid $\mathcal{G}(n)$. Thus, $R^{n_k}(u) \geq R^{n_{k+1}}(u)$ for each k. Since the sequence $\{R^{n_k}, k = 0, 1, 2, \ldots, \}$ is bounded and decreasing, it converges to the limit $R(u)$ of its sub-sequence $\{R^{n_k(l)}\}$. This concludes the proof. □

The chessboard based $R(u)$-estimating procedure involves two computational steps. The first is performed once at the beginning of the estimation procedure and computes the collection of indices t that make up $\mathcal{S}(u)$. This determines the collection of indices $J^n(u)$ for each n. The second involves iteratively solving the LP (21) for larger values of n.

The collection $\mathcal{S}(u)$ is found by exploring the 2^d combinations of the type in (20), but this computation can

potentially take an exponential (in d) number of steps in the worst case scenario. The computation can however be speeded up for many cases, and we expect it to perform reasonably on average. First note that only distinct values of v_i need be considered: the indices of form $t = (t_1, \ldots, t_i, \ldots, t_d)$ included in $\mathcal{S}(u)$ will be identical for obligors k and l if their default losses satisfied $v_k = v_l$. The computation can also be made less intensive by ordering all obligors i by their (unique) default sizes v_i and then exploring the possible combinations in (20) in a depth-first fashion.

The second major computational step consists of solving the chessboard LP (21) iteratively for increasing n to obtain increasingly accurate estimates of $R(u)$. Theorem 3.1 shows that the estimates provided are always larger than the true maximal value. Thus, a conservative estimate of the worst-case value-at-risk is obtained from the approximation for any n. These LPs have an n^d number of variables, and thus the size of the LP can grow fast with n. One would ideally like to avoid iteratively solving the LP (21) by being able to choose a value of n that will yield a sufficiently accurate estimate of $R(u)$. The chessboard distributions constructed by the LP of size n might match the covariance Σ only approximately (i.e., only satisfy (12)), but are within a calculable bound from Σ (refer to Section 2.1). Thus, a result on the nature of $R(u)$ with regards to the Σ values (continuity, differentiability, etc.) would help predict an appropriate value of n. We are working towards such results, but have no such to present as of this writing.

4 HANDLING LARGE PROBLEMS

The linear program based estimation procedure of Section 3 is increasingly harder to run with the parameter d, since the sizes of these linear programs grow very quickly with d. So, solving LPs in their naive formulation for large portfolios typically with hundreds of obligors will not be possible. One approach to get around this problem is to reformulate the LP (21) as an axially symmetric multi-index transportation problem (Queyranne and Spieksma 2001) with additional constraints, and then leverage this specialized structure to provide high quality approximations of the optimal objective value. The approximation procedures available in the literature (Queyranne and Spieksma 1997) have to be tweaked to handle the extra constraints in (12) and we are actively pursuing this avenue.

Another aspect of these large-portfolio problems provides some respite from this curse of dimensionality. Portfolios are often analysed by first identifying and grouping together *homogeneous* obligors with very similar characteristics; see Glasserman and Li (2005) and Bassamboo, Juneja, and Zeevi (2005). The aggregation is performed for instance by classifying all obligors that belong to the same credit rating in an industry tracking index as homogeneous. Suppose K such groupings are formed. As per this assumption, we shall define:

Definition 4.1 *Homogeneous obligors*

1. *Follow the same marginal distributions F_k for each group $k = 1, \ldots, K$,*
2. *Default with the same probability p_k and thus share the same default threshold v_k,*
3. *Induce an identical loss e_k for the creditor if they default,*
4. *Are correlated with value $\sigma(k)$ with fellow homogeneous obligors in the group k, and*
5. *Are correlated with value $\sigma(k,l)$ with each obligor in group j, $\forall\, k, j = 1, \ldots, K$.*

This has many implications on the structure of the estimation procedure. First, since the marginals and the default thresholds are the same for homogeneous obligors, the $\mathcal{S}(u)$ enumeration step essentially becomes a search over K values of v_k, reducing the complexity from being an exponent of d to that of K (presumably $<< d$).

Second, all homogeneous obligors have symmetric correlation relations, and thus the correlation matrix has large blocks with identical values. Since the exact ordering of the homogeneous obligors within the group does not change the formulation (21) we need consider only feasible solutions q that are symmetric with respect to each set of homogeneous obligors.

To see why, consider the $3-d$ case where variables X_2 and X_3 share a homogeneous correlation structure. Suppose q' is an optimal solution to the LP (21). Then an alternate optimal solution q'' can be obtained by simply re-ordering each $q''(j_1, j_2, j_3) = q'(j_1, j_3, j_2)$. In other words, a symmetric solution $\bar{q} = (q' + q'')/2$ can also be identified for each optimal q'.

Thus, one can set large blocks of the variables q to be equal in value, which reduces the effective dimension of the LP (21) from d to K. This homogenization-based simplified procedure is the one tested in the results in Section 6.

5 OTHER CREDIT RISK MEASURES

The value-at-risk measure treated in Section 3 is insensitive to the magnitude of loss incurred due to mass defaults. Certain other credit-risk measures have been proposed in the literature to provide an idea of this magnitude. Chief amongst these is the *expected shortfall* metric (Bassamboo et al. 2005) which weighs large losses by their magnitude. It is defined as the expected excess loss conditioned on the event that the loss exceeds a large threshold. Suppose the loss L associated with a credit portfolio is modeled as in (2). Then, the expected shortfall at threshold u is $E[L-u|L>u]$.

This can be re-written as

$$E[L-u|L>u] = \frac{E[(L-u)I(L>u)]}{P(L>u)} = \frac{E[(L-u)^+]}{P(L>u)},$$

where $(x)^+ = \max\{x,0\}$. The denominator is the familiar value-at-risk measure we have already treated in Section 3.

The numerator represents the expected loss in excess of the threshold u and is arguably an interesting statistic on its own. It is also a linear function of the qs of our formulation. Suppose the loss L takes the (fixed) value L_t over each S_t in $\mathcal{S}(u)$. Then

$$E[(L-u)^+] = \sum_{t:S_t \subseteq \mathcal{S}} L_t \sum_{J_t^n(u)} q(j).$$

This form is very similar to the objective function of the LP (21), and in fact has the coefficients 1 replaced with appropriate L_t. One can thus use the same LP formulation (21) to maximize (minimize) the $E[(L-u)^+]$ value to find accurate estimates of the true maximum (minimum) over the collection $\mathcal{F}$.

Similar estimates can be found for the *minimum* value-at-risk. These can then be combined to obtain bounds of the form

$$\frac{E_{min}[(L-u)^+]}{P_{max}(L>u)} \leq E[L-u|L>u] \leq \frac{E_{max}[(L-u)^+]}{P_{min}(L>u)} \quad (23)$$

Theorem 3.1 shows that the estimates provided by the maximization (minimization) formulation are always larger (smaller) than the true maximal (minimal) value. Thus, valid bounds are obtained from the approximations for any n.

Note that the true P_{min} can in theory be 0, but in most cases of interest this can reasonably be expected not to be the outcome. For instance when the risk of large losses beyond a high threshold u is being evaluated, and individual obligors are durable enough to default only for low threshold values v_i we can expect the interesting subset $\mathcal{S}(u)$ of the support of the joint distribution to contain the non-zero-mass support of entire sub-intervals of certain marginals. This then implies that P_{min} has a value greater than or equal to the mass assigned by that marginal to the relevant sub-interval, and thus is non-zero.

6 NUMERICAL EXPERIMENTS

These numerical experiments are performed to test the quality of the estimates provided for the maximal value-at-risk defined in (1) by methods that assume a multivariate *normal*- or $t-$copula structure for the latent variables X_i underlying the loss function L as defined in (2). The portfolio sizes d are varied from 100 to 500. We start with $K = 1$ homogeneous group of obligors and study how the estimate quality changes as K is increased to 10. The homogeneous groups are chosen with appropriately differing parameters. The results will be presented at the talk accompanying this article.

REFERENCES

Basel Committee on Banking Supervision, 2002. The new basel capital accord. <http://www.bis.org/bcbs/bcbscp2.htm>.

Bassamboo, A., S. Juneja, and A. Zeevi. 2005. Expected shortfall in credit portfolios with extremal dependence. In *Proceedings of the 2005 Winter Simulation Conference*, ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines.

Bassamboo, A., S. Juneja, and A. Zeevi. 2007. Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research*. To Appear.

Billingsley, P. 1986. *Probability and measure*. 2nd ed. New York: Wiley.

Billingsley, P. 1995. *Probability and measure*. 3rd ed. New York: Wiley.

Ghosh, S. 2004. Dependence in stochastic simulation models. Ph. D. thesis, Department of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

Ghosh, S., and S. G. Henderson. 2002. Chessboard distributions and random vectors with specified marginals and covariance matrix. *Operations Research* 50 (5): 820–834.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2002. Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance* 12:239–269.

Glasserman, P., and J. Li. 2005. Importance sampling for portfolio credit risk. *Management Science* 51:1643–1656.

Gupta, G., C. Finger, and M. Bhatia. 1997. Credit metrics technical document. Technical report, J.P. Morgan & Co., New York.

Kealhofer, S., and J. Bohn. 2001. Portfolio management of credit risk. Technical report, KMV working paper, New York.

Mashal, R., and A. Zeevi. 2003. Beyond correlation: extreme co-movements between financial assets. <http://ssrn.com/abstract=317122>.

Queyranne, M., and F. C. R. Spieksma. 1997. Approximation algorithms for multi-index transportation problems with decomposable costs. *Discrete Applied Mathematics* 76:239–254.

Queyranne, M., and F. C. R. Spieksma. 2001. Multi-index transportation problems. In *The Encyclopedia of Opti-*

mization, ed. F. C. and P. Pardalos, 450–456. Dordrecht: Kluwer.

AUTHOR BIOGRAPHIES

SOUMYADIP GHOSH is a Research Staff Member of the Mathematical Sciences Department at IBM T.J. Watson Research Center, Yorktown Heights, NY. His research interests encompass simulation theory and practice, in particular on input dependence modeling and optimization. He can be contacted at <ghoshs@us.ibm.com>.

SANDEEP JUNEJA is a faculty member in the School of Technology and Computer Science at the Tata Institute of Fundamental Research. His research interests include Monte-Carlo methods for stochastic systems. His email address is <juneja@tifr.res.in>, and his web page is <www.tcs.tifr.res.in/~sandeepj>.